Structure 14, 401–414, March 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2005.11.022

Structural Model of the mAb 806-EGFR Complex Using Computational Docking followed by Computational and Experimental Mutagenesis

Arvind Sivasubramanian,[1,4] Ginger Chao,[2] Heather M. Pressler,[2] K. Dane Wittrup,[2] and Jeffrey J. Gray[1,3,4,*]

[1] Department of Chemical & Biomolecular Engineering
Johns Hopkins University
3400 North Charles Street
Baltimore, Maryland 21218

[2] Department of Chemical Engineering and Biological Engineering Division
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

[3] Program in Molecular & Computational Biophysics
Johns Hopkins University
Baltimore, Maryland 21218

[4] Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
Baltimore, Maryland 21231

Summary

In this work, we combined computational protein-protein docking with computational and experimental mutagenesis to predict the structure of the complex formed by monoclonal antibody 806 (mAb 806) and the epidermal growth factor receptor (EGFR). We docked mAb 806, an antitumor antibody, to its epitope of EGFR residues 287–302. Potential mAb 806-EGFR orientations were generated, and computational mutagenesis was used to filter them according to their agreement with experimental mutagenesis data. Further computational mutagenesis suggested additional mutations, which were tested to arrive at a final structure that was most consistent with experimental mutagenesis data. We propose that this is the EGFR-mAb 806 structure, in which mAb 806 binds to an untethered form of the receptor, consistent with published experimental results. The steric hindrance created by the antibody near the EGFR dimer interface interferes with receptor dimerization, and we postulate this as the structural origin for the antitumor effect of mAb 806.

Introduction

Excessive signaling from epidermal growth factor receptor (EGFR) and homolog ErbB2 has been implicated (Yarden and Sliwkowski, 2001) in a broad spectrum of epithelial carcinomas, including breast, ovarian, head and neck, lung, pancreatic, and colorectal cancer. Antibody targeting of the extracellular region of these receptors is known to mitigate the effect of aberrant signaling and to induce an antitumor effect (Harari, 2004; Sridhar et al., 2003). The crystal structures of the antitumor antibodies Trastuzumab (Cho et al., 2003), Cetuximab (Li et al., 2005), and Pertuzumab (Franklin et al., 2004) bound to their respective receptor antigens (ErbB2, EGFR, and ErbB2) have been solved. These structures have aided in the elucidation of the mechanisms of action of these antibodies.

Crystal structures of the EGFR extracellular domain in the monomeric (Protein Data Bank [PDB] identifiers 1NQL, 1YY9) and dimeric states complexed with ligands such as TGF-α (1MOX) (Garrett et al., 2002) and EGF (1IVO) (Ogiso et al., 2002) have been reported. The monomeric EGFR is autoinhibited due to intramolecular tethering interactions between the domain II dimerization arm and a domain IV β loop (Ferguson et al., 2003). Figure 1 depicts the EGFR conformation change from the tethered monomer to a hypothetical untethered state seen in the dimer, which leads to the formation of the dimer. The hypothetical extended monomer state is reached via a 130° counterclockwise rotation of the rigid body comprised by domains I and II that releases the intramolecular tether in the tethered conformation (1NQL) and brings domains I and III closer to each other. This motion frees the dimerization arm to interact with its counterpart on a second EGFR to form the active EGFR dimer. While it is possible that ligand binding induces the EGFR conformational change, the current view is that the ligand stabilizes the extended conformation from an ensemble of states sampled by the flexible and dynamic EGF receptor (Walker et al., 2004).

The antitumor antibody mAb 806 was first described by Mishima and coworkers (Mishima et al., 2001). mAb 806 binds to the extracellular region of the EGFR when it is either truncated in EGFRvIII or overexpressed, indications prevalent in several types of cancers; however, mAb 806 does not bind the endogenous EGFR in tissues such as liver and skin (Johns et al., 2002; Luwor et al., 2001; Mishima et al., 2001). Previous experimental work with immunoblotting and yeast surface display of EGFR fragments identified a disulfide bonded loop (amino acids 287–302) in EGFR domain II that contains the complete mAb 806 epitope (Johns et al., 2004). Subsequently, individual residues in the peptide that are energetically important in mAb 806 binding were identified by using random mutagenesis and yeast surface display (Chao et al., 2004). These residues appear to be accessible if residues 6–273 are removed, as they are in EGFRvIII, but inaccessible in the wild-type tethered monomer conformation. This is consistent with experimental data that mAb 806 does not bind the receptor in solution, but binds when the receptor is either immobilized on ELISA plates or denatured by heat (Johns et al., 2002). To explain these observations, it has been suggested that a putative intermediate state along the EGFR conformational transition pathway exists that exposes the mAb 806 epitope residues, thus leading to stronger antibody binding (Johns et al., 2004). However, it is not clear if the intermediate state is closer to the tethered or extended receptor monomer conformation.

Experiments on mAb 806 binding to various mammalian cell lines tend to confirm the notion of an "intermediate" state. mAb 806 binds robustly to EGFR mutants with a reduced tethering capacity through either deletion of the dimerization arm or point mutation (Walker

*Correspondence: jgray@jhu.edu

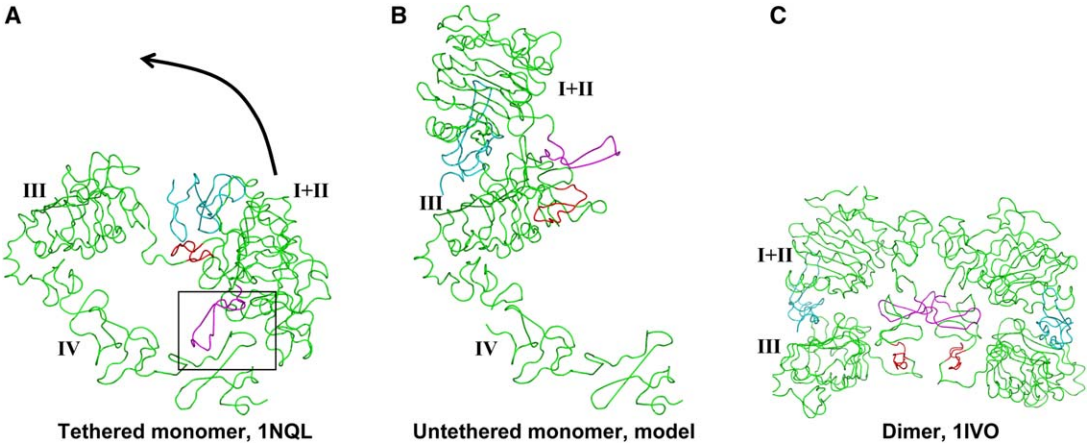

Figure 1. EGFR Conformational Change

(A–C) The dimerization arm, the mAb 806-peptide epitope, and the EGF ligand are colored magenta, red, and cyan, respectively. EGFR domains are labeled. The intramolecular tether in (A) is highlighted by the box. Upon release of the tether, EGFR in the tethered state ([A], 1NQL) undergoes a conformational change to a hypothetical untethered, extended monomer whose coordinates are taken from chain A in the dimer (1IVO) for domains I–III and from the tethered monomer (1NQL) for domain IV since this domain is disordered in the dimer. The arrow shows the direction in which EGFR domains I and II move to reach (B) from (A). The dimer ([C], 1IVO) is formed by interaction of two receptors in the extended state. EGFR domain IV is not shown in (C). The dimerization arm is proximal to the peptide epitope recognized by mAb 806.

et al., 2004). mAb 806 binds poorly to the parental U87MG cell line that endogenously expresses the wild-type EGFR ($\sim 10^5$ receptors per cell) in the absence of EGFR gene amplification (Johns et al., 2002). However, in the A431 cell line where the EGFR is overexpressed (2×10^6 receptors per cell) due to gene amplification, mAb 806 recognizes about 10% of the total EGFR molecules (Johns et al., 2002). Recent experiments (Johns et al., 2005) show that mAb 806 preferentially binds to immature high-mannose forms of the wild-type EGFR. When the EGFR is overexpressed, such immature EGFR forms appear on the cell surface, possibly due to saturation of specific glycosylation pathways. The altered glycosylation state in this immature receptor may promote receptor untethering, thus leading to increased mAb 806 binding. This appears to explain why mAb 806 does not bind wild-type EGFR, but binds a fraction of the cell-surface EGFR when it is overexpressed.

While there is significant biochemical and cellular information on mAb 806 binding to EGFR, no molecular structure has been determined for the mAb 806-EGFR complex. X-ray crystallization, while providing complete and high-resolution information, is time consuming and can be limited by the ability of the complex to form a crystal. This particular complex is also too large for accurate determination by standard NMR techniques. An inexpensive and rapid alternative is to use protein structure prediction tools to create the antibody structure and computational docking to predict the complex model. While these tools involve considerable uncertainty, in the recent international blind prediction challenge, the Critical Assessment of Predicted Interactions or CAPRI, several docking algorithms (Aloy et al., 2001; Berchanski et al., 2004; Camacho and Gatchell, 2003; Comeau et al., 2004; Dominguez et al., 2003; Fernandez-Recio et al., 2003; Gray et al., 2003b; Li et al., 2003; Mandell et al., 2001; Ritchie and Kemp, 2000; Schneidman-Duhovny et al., 2003; Smith et al., 2005; Wang et al., 2005; Zacharias, 2003) have been tested, and many (Berchanski et al., 2004; Camacho and Gatchell, 2003; Dominguez et al., 2003; Fernandez-Recio et al., 2003; Gray et al., 2003b; Li et al., 2003; Mandell et al., 2001; Wang et al., 2005) have produced high-quality predictions for multiple protein-protein targets in CAPRI rounds 3–5.

Our current target involves two significant modeling challenges that have not been previously explored or have been found to be difficult. First, the structure of one of the components, mAb 806, is not itself known and must be predicted with homology modeling. Second, the EGFR is quite large and exhibits significant domain flexibility. However, by incorporating biochemical information available on mAb 806 and EGFR, and by iteratively testing docking predictions by experimental mutagenesis, the accuracy of docking predictions can be improved to the point of reasonable confidence in the final result. Experimental data about the binding interface are commonly employed in docking studies (for a recent review, see van Dijk et al. [2005]), and here we highlight only a few representative examples. In single-mutagenesis studies, the change in binding affinity is measured after mutation of a particular residue. Loss of binding after mutation at a particular position suggests, but does not guarantee, that the mutated residue is located at the binding interface. Neutral binding affinity change at a position does not necessarily suggest that a residue is not located at the interface, as it may be possible that it simply does not contribute much to the binding energy. These cases are observed in the interaction of a heterotrimeric G protein subunit with other subunits and a transmembrane receptor studied with affinity tests of 100 mutants of the central subunit (Onrust et al., 1997). When a method assumes that nondisruptive mutation positions are explicitly not at the protein-protein interface and disruptive or partially disruptive positions are, the models can be very sensitive to the experimental information (Herzyk and Hubbard, 1998). Appropriately, HADDOCK is one

notable docking program that incorporates experimental information in an ambiguous fashion, allowing for the use of NMR chemical shift perturbations and residual dipolar couplings in addition to mutagenesis data (Dominguez et al., 2003). Mutagenesis can also be used subsequent to docking for validation. A complex of tissue factor, factor VIIa, and factor Xa constructed by docking allowed for rational selection of target residues for site-directed mutagenesis that were in agreement with the model (Norledge et al., 2003). In a study with a target similar to the current work, a homology model of a recombinant antibody was docked to the α-bungarotoxin and subsequently validated by mutations suggested by human inspection of the two putative binding models (Bracci et al., 2003). Massova and Kollman (1999) developed a model for "computational alanine scanning" by using a molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA)-based energy function and applied it to study the interactions of the oncoprotein Mdm2 to the N-terminal stretch of tumor suppressor protein p53. Recently, explicit energy calculations on an ensemble of mutated sequences were exploited to improve predictions of the folding of lattice proteins (Nanda and DeGrado, 2005). Similarly for protein-protein docking model structures, more information might be gleaned from mutagenesis experiments by exploiting energy calculations for the specific mutations.

Our objective is to predict the mAb 806-EGFR complex structure by using a protocol that combines the protein docking program RosettaDock (Gray et al., 2003b) with computational (Kortemme and Baker, 2002) and experimental mutagenesis. The RosettaDock method employs a low-resolution rigid-body Monte-Carlo search followed by an all-atom, simultaneous optimization of backbone displacement and backbone-dependent side chain rotamer conformations by using Monte-Carlo minimization. RosettaDock and the RosettaInterface mutagenesis program (Kortemme and Baker, 2002) both capture side chain flexibility and employ a comprehensive energy function dominated by van der Waals interactions, solvation, and hydrogen bonding. The performance of the RosettaDock algorithm has been successfully tested (Daily et al., 2005; Gray et al., 2003a) in several rounds of CAPRI, and RosettaDock correctly predicted the complex structure of dockerin and cohesin in CAPRI round 4 by using a homology model with considerable structural errors (Daily et al., 2005). Here, we propose to use the available mutagenesis data on mAb 806 binding to EGFR (Chao et al., 2004) to not only identify interface residues, but to verify via computational mutagenesis (Kortemme and Baker, 2002) if a mutation produces similar outcomes in a docking decoy and in experiments. To the best of our knowledge, this is the first instance in which computational and experimental mutagenesis are iteratively combined with docking to isolate native-like decoys from other incorrect docking solutions.

Results

Figure 2 outlines our overall protocol for the determination of the mAb 806-EGFR complex structure. Experimental results support the hypothesis that mAb 806 binds a transitional form of the EGFR, but the exact location of this conformation along the path from tethered monomer to hypothetical extended monomer is unknown and its structure has not been determined. Therefore, we use the experimentally determined peptide epitope, instead of the full EGFR protein, as the docking partner for mAb 806. We independently docked mAb 806 to five crystal structure conformations of the EGFR peptide (residues 287–302), which contains the mAb 806 epitope as demonstrated by previous experiments. Since a crystal structure is also not available for mAb 806, we built a homology model by using the Web Antibody Modeling (WAM) tool (Whitelegg and Rees, 2000). We use the RosettaDock algorithm to generate decoys, or potential mAb 806-EGFR structures, and rank them by using an all-atom energy function. Next, we perform computational mutagenesis on low-energy decoys to identify mAb 806-peptide orientations in conformity with published experimental mutagenesis data (Chao et al., 2004). We structurally superpose the models with the EGFR ectodomain crystal structure to eliminate unphysical orientations of the antibody that arise from docking mAb 806 to the peptide epitope without considering the steric hindrance due to the nonpeptide EGFR residues. In this manner, we identify three potential models, and additional computational mutagenesis results suggest new aromatic residue substitutions to distinguish between these models. We select the final mAb 806-EGFR complex model as the one most consistent with these new experimental results.

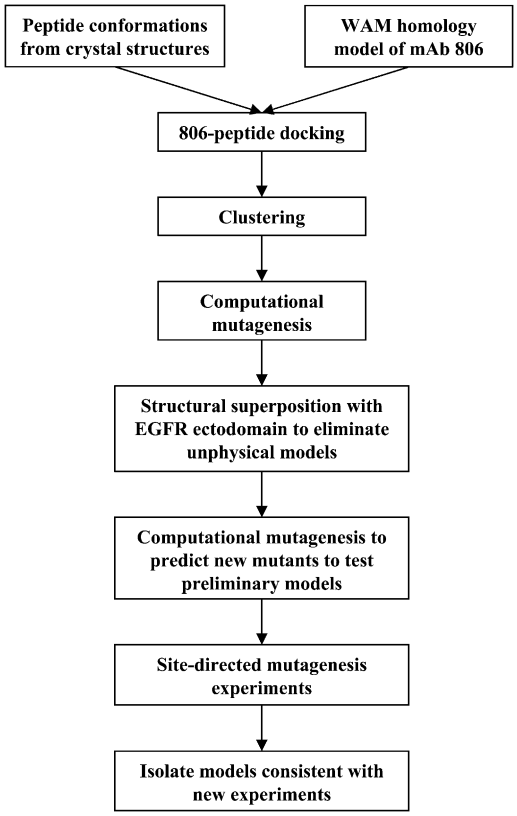

Figure 2. Protocol for the Prediction of the mAb 806-EGFR Complex Structure

Differences in the Peptide Epitope Backbone Conformation between the Tethered and the Untethered EGFR Crystal Structures

Figure 3 shows the mAb 806-peptide epitope (amino acid sequence CADSYEMEEDGVRKC) backbone conformations from the dimer crystal structures (1IVO, 1MOX) superposed onto that of the tethered monomer (1NQL). The pairwise C_α rmsd between the different conformations is within 2 Å (Table 1) for the epitope residues and 1–1.5 Å for the larger EGFR segment 275–305 (data not shown). There is a tight consensus between the independently determined peptide backbones in the N- and C-terminal regions (287–291 and 299–302); variation in the backbone occurs between residues 292 and 298. The consensus between these conformations suggests an overall stability of the peptide due to the disulfide bond between the N- and the C-terminal cysteines. On the other hand, the availability of different backbone conformations is fortunate because it allows for the use of different starting structures in the docking simulations to capture the variability inherent in the exposed peptide epitope.

A closer look at the peptide conformation from the tethered monomer reveals that the C_β of residue E293, a known hotspot for mAb 806 binding, is oriented differently relative to the dimer crystal structures (Figure 3A). We term the backbone conformation of the peptide in 1NQL as the "flipped out" conformation, as the side chain of E293 is exposed at the surface of the tethered monomer. On the other hand, the peptide conformations in 1MOX and 1IVO are "flipped in," as the E293 side chain points toward the protein interior and is inaccessible to antibody. The average pairwise angle between the C_α-C_β vectors at position 293 from the various dimer crystal structures is 15.4°, indicating that the side chains have similar directionality. In contrast, the average pairwise angle between the C_α-C_β vector at position 293 from the 1NQL peptide and those from the dimer peptide conformations is 130.5°, signaling a complete reversal of the C_β directionality. Although only the C_β coordinate for E293 is resolved in 1NQL, the orientation change is supported by the corresponding change in the directionality of the bulky Y292, which is clearly resolved in all of the structures, and M294, which has only a resolved C_β in 1NQL (Figure 3B). For the dimers alone, the orientation of the C_α-C_β vectors differ on average by 34.6° and 23.6° for positions 292 and 294, respectively. However, the C_α-C_β vectors of residues 292 and 294 in the 1NQL peptide change directions by 84.4° and 91.6°, respectively, on average from the dimer structures. By the same measure, the remaining epitope side chains are nominally oriented in the same direction in both the dimer and the monomer crystal structures. Thus, the crankshaft-type move involving residues 292–294 essentially reverses the directionality of these side chains without affecting the remaining peptide residues. It is unclear if the crankshaft move is due to the dynamics of the peptide epitope in solution or if it reflects conditions of the crystallization process.

mAb 806 Is Predicted to Adopt a Predominantly Canonical Structure

Potential uncertainties in the WAM homology model are restricted to the antigen binding CDR (complementarity-determining region) loops, as the antibody framework is highly conserved (Chothia and Lesk, 1987). In general, all CDR loops except H3 can be assigned to canonical classes that depend on the identity of key residues in the antibody sequence (Al-Lazikani et al., 1997; Chothia and Lesk, 1987). A comparative analysis of the main chain CDR conformations (excluding CDR H3) in high-resolution antibody structures revealed that 71 out of 79 CDR loops had a crystal structure conformation very close to the standard conformation of the respective canonical classes (Al-Lazikani et al., 1997). Of the remaining eight, six differ in the orientations of a single peptide group, and two in CDR H2 involve nonstandard conformations that are allowed by the presence of three glycine residues in a surface loop.

With the exception of a few residues, the amino acid sequences for mAb 806 CDRs L1–L3, H1, and H2 are nearly identical to the corresponding sequences in antibody M02/05/01 (PDB 1C12, resolution 2.6 Å). These five CDRs are assigned (http://antibody.bath.ac.uk/autoalign.html) identical canonical conformations (2, 1, 1, 5, and 1) in both antibodies. These CDR backbone conformations in the homology model are essentially identical to the respective conformations in PDB 1C12. MAb 806 has a 7 residue CDR H3 (AGRGFPY). This is a short loop with a small conformational space and few reasonable possibilities for placement. The closest homolog (GGTGFPY) is found in CDR H3 of antibody PC282 (PDB 1KCV, resolution 1.8 Å), which also has CDRs H1 and H2 that are homologous with mAb 806. The CDR H3 regions in the homology model and PDB 1KCV align with a C_α rmsd of 1.2 Å, suggesting that WAM accurately models the local conformation of CDR H3. However, with a global alignment of both the light and heavy chains, the C_α rmsd for the CDR H3 residues increases to 5.3 Å, even though the rmsd for the antibody as a whole is ~1 Å. This indicates that the positioning of the H3 loop relative to the framework is altered between mAb 806 and 1C12. This alteration might originate from the differences in the sequence of the H3 itself or the N-terminal stem region (CVT versus CAR for PDB 1KCV) or from variations in the packing at the light-heavy chain interface.

mAb 806-Peptide Docking, Mutagenesis, and Structural Superposition

The RosettaDock docking algorithm captures side chain flexibility by using a backbone-dependent rotamer library, but it does not model protein backbone flexibility. To allow for variability in the peptide backbone conformation, we used RosettaDock to independently dock mAb 806 to each of the five peptide backbone conformations from the EGFR crystal structures, generating 10^5 decoys in each simulation. We isolated the top 2000 decoys by energy from each run and clustered these by using a hierarchical clustering algorithm with a tight clustering radius of 0.5 Å. This resulted in $O(100)$ clusters from each run. Each cluster is represented by the decoy with the lowest energy in the cluster, and the cluster size is the number of decoys in the given cluster. Clusters are numbered in order of decreasing cluster size.

In the next stage of our protocol, we used the RosettaInterface computational mutagenesis approach (Kortemme and Baker, 2002) to identify docking clusters

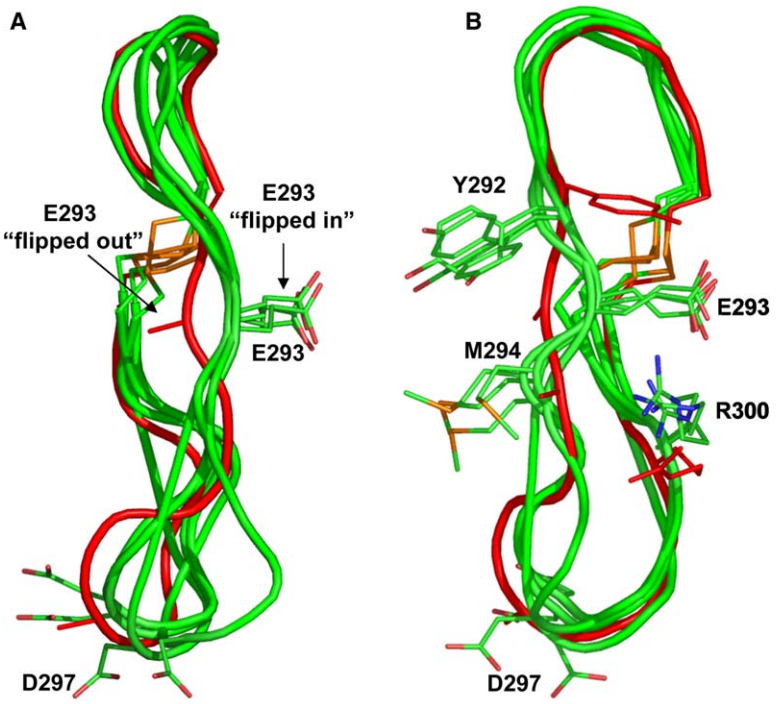

Figure 3. Peptide Epitope Backbone from EGFR Crystal Structures

(A and B) Superposition of the mAb 806-peptide epitope backbone from different EGFR crystal structures. (A) Change in E293 side chain directionality; (B) alternate view additionally showing side chains Y292, M294, and R300. The "flipped out" backbone and side chains from the tethered monomer (1NQL) are shown in red; the "flipped in" backbone and side chains from the dimers are shown in green and CPK colors, respectively; and the disulfide bond is shown in orange. Side chains or β carbons, as available in the crystal structure, are shown for labeled residues.

that were consistent with experimental data for the mAb 806-EGFR system. The RosettaInterface algorithm employs the Rosetta energy function to estimate the change in the binding free energy ($\Delta\Delta G$) of a protein-protein complex due to residue mutations. We retained the hard cutoff of $\Delta\Delta G_{calc} > 1.0$ kcal/mol to qualitatively identify hotspot residue mutations, since Kortemme and Baker (2002) found that this criterion was sufficient to correctly identify 80% of hotspot residue mutations in a set of 19 high-resolution protein-protein complex crystal structures. $-1.0 < \Delta\Delta G_{calc} < 1.0$ indicates a neutral mutation, while $\Delta\Delta G_{calc} < -1.0$ indicates a mutation that potentially increases binding affinity. Experimental data on the mAb 806-EGFR system has been published by Chao and coworkers, who used random mutagenesis and yeast surface display to identify individual residues in the peptide that are energetically important in mAb 806 binding (Chao et al., 2004). It was shown that four peptide residue mutations (E293K, G298D, G298S, V299D) result in a large binding loss, with a $\Delta\Delta G > 5$ kcal/mol. Six mutations (E293A, E293D, E293G, D297Y, G298A, K301E) result in partial binding loss, with $\Delta\Delta G \sim 2$ kcal/mol. Thirteen EGFR mutants (G288A, A289K, D290A, S291A, Y292A, M294A, E295A, E296A, D297A, V299K, V299A, R300A, and K301A) retain wild-type binding. The data suggest that residue E293 is particularly important for mAb 806 binding, as conservative (E293D), nonconservative (E293K), and alanine mutations all result in binding loss. Computational mutagenesis was used to predict the effect of these 23 residue mutations in the decoy clusters obtained from docking. The computational results were compared to the experimental mutagenesis results, and we refer to a match between the two results as a "hit." We then filtered the clusters by stipulating at least 5 hits for the 10 mutations that are known to produce at least some binding loss experimentally, and 10 hits for the 13 mutations that retained wild-type binding. The accuracy of the computational mutagenesis technique does not allow us to exactly match the experimental $\Delta\Delta G$ values or draw distinctions between experimental mutations that cause partial or large binding loss. After application of the mutagenesis filter, 35 clusters remained that were broadly consistent with the experimental mutagenesis data.

Next, we structurally superimposed docking clusters that cleared the mutagenesis filter with the tethered monomer to identify and eliminate clusters with mAb 806-peptide orientations that are unphysical in the context of the full-length EGFR. These unphysical orientations arose from docking mAb 806 to the peptide epitope without considering any steric hindrance due to the nonpeptide EGFR residues. Since mAb 806 does not bind the tethered EGFR in experiments, the superposition could legitimately result in clashes of mAb 806 residues with the nonpeptide EGFR residues. Therefore, decoys were eliminated only if they contained severe clashes of the antibody with EGFR residues that are not predicted to move relative to the peptide during the conformational change from tethered monomer to extended dimer. By the same token, we retained decoys that had minor clashes of less than five amino acids with the EGFR ectodomain, since antibody binding may relieve such clashes via an induced fit mechanism. Thus, the superposition filter eliminates most of the physically unrealistic mAb 806-peptide orientations without being overly restrictive regarding allowable orientations at this stage.

At the end of this procedure, three clusters of the mAb 806-EGFR peptide complex structure remained and were labeled as models 1, 2, and 3. Table 2 lists the computationally calculated $\Delta\Delta G$ values for these models for the EGFR residue mutations that lead to loss of mAb 806 binding. The peptide backbones in these models come

Table 1. Peptide Backbone Conformational Differences

	1IVO:B	1MOX:A	1MOX:B	1YY9	1NQL
1IVO:A	0.94	1.76	0.84	0.85	1.10
1IVO:B		1.93	1.27	1.13	1.41
1MOX:A			1.49	1.69	1.69
1MOX:B				0.63	1.03
1YY9					0.95

Pairwise Backbone C_α Rmsd Values for Peptide Conformations from EGFR Crystal Structures, Residues 287–302. Data are also shown for the peptide from a recently published EGFR crystal structure (PDB 1YY9), which was not used in the docking simulations.

from two different EGFR crystal structures. Models 1 and 2 use the "flipped in" peptide conformation from the dimer (1IVO chain A), while model 3 uses the "flipped out" backbone from the tethered monomer (1NQL). The three models have 5, 6, and 6 hits, respectively, suggesting qualitative agreement between computational and experimental mutagenesis for the binding loss mutations. Model 1 has 11 of 13 hits for the neutral mutations, while models 2 and 3 produce 13 of 13 hits (data not shown).

Apart from these three clusters, ten other clusters that produce at least four of ten binding loss hits also satisfy the structural superposition criterion. However, we eliminated these clusters for either of two reasons: the cluster either had no binding loss hits for any of the important E293 mutations, or it produced hits for only the E293 mutations and no hits for any of the other mutations.

Interface Side Chain Interactions in the Preliminary Models

Figure 4 shows the mAb 806-peptide orientations and the positions of key side chains in the three models, and Table 3 shows the residue-level interface contacts for peptide residues 293–301. In all models, the peptide residues interact primarily with residues from CDRs L3 and H2. In model 1, all peptide residues except V299, K301, and C302 contact the mAb 806 CDRs. In model 2, residues 288–291, 297, 299, 301, and 302 do not contact the antibody. Model 3 interactions are different, with all residues except 287−289 and 300 involved in interface contacts. In model 1, E293 contacts residues in CDRs L2 and H3. In models 2 and 3, E293 makes similar contacts with CDR H2 residues. D297 makes one and zero contacts in models 1 and 2, respectively, while, in model 3, D297 is buried in a pocket formed by 5 residues from CDRs L2, L3, and H3. Interestingly, models 1 and 2 are similar in global arrangement, but model 3 differs from either of them. Model 2 may be obtained from model 1 by rotating the peptide about the axis shown in Figure 4A.

Position of mAb 806 with Respect to the Full-Length EGFR Conformation Obtained via Superposition

The presence of a hypothetical "intermediate" state formed by receptor untethering has been postulated (Johns et al., 2004) based on the observation that mAb 806 does not bind the tethered EGFR. Apart from this, it is not clear if this "intermediate" state is close to the tethered EGFR, or if it resembles the hypothetical extended monomer in Figure 1B. In Figure 5, we show the relative position of mAb 806 and the full-length EGFR in the three models after structural superposition with the tethered (Figures 5A–5C) and the hypothetical extended monomer states (Figures 5D–5F). In the superposition of model 1 with the tethered EGFR, the antibody light chain has minor clashes (but not interpenetration) with residue fragment 261–286 and also parts of domain III. Superposition of model 2 with the tethered EGFR is similar to model 1, except that the heavy chain is involved in the clashes. In model 3, the antibody light chain entirely interpenetrates into the EGFR dimerization arm, and clashes (not seen in the figure) of the heavy chain with parts of EGFR domain IV also occur.

The situation is different for the structural superposition with the hypothetical extended monomer. For models 1 and 2, there is complete interpenetration of the antibody and EGFR domain III. For model 3, the EGFR transition from the tethered to the untethered but nondimer state eliminates most of the interpenetration, and visual inspection confirms that only some minor clashes remain. However, there are also new minor clashes between the heavy chain and EGFR residues that occur distal to the peptide.

These observations indicate that the location of the intermediate state is close to the tethered conformation if model 1 or 2 is correct. The minor clashes of models 1 and 2 with the tethered state can be relieved via local movement of domains I, II, and III relative to the peptide. The required movements are within the range of the observed EGFR conformational change between the untethered monomer and extended dimer crystal structures (Experimental Procedures). However, due to the severe clashes with domain III in the hypothetical extended monomer, the mAb 806-peptide orientation in either of these models does not support an intermediate close to this EGFR conformation. If model 3 were the correct binding mode, the superposition indicates an "intermediate" state close to the hypothetical extended

Table 2. Initial Computational Mutagenesis Filter

Model Number	Cluster Number	Crystal Structure	Cluster Size	E293D (+)	E293K (−)	E293A (+)	E293G (+)	D297Y (+)	G298A (+)	G298S (−)	G298D (−)	V299D (−)	K301E (+)	Hits
1	10	1IVO:A	12	0.6	**1.2**	0.7	**1.0**	0.0	**2.5**	**2.3**	**3.0**	0.0	0.0	5
2	53	1IVO:A	3	**1.3**	**1.8**	0.9	**1.2**	−0.1	**1.8**	**2.2**	**4.9**	0.0	0.3	6
3	11	1NQL:A	27	**1.3**	**1.5**	**1.3**	**1.4**	**40.3**	0.4	0.2	**1.0**	0.0	0.4	6

$\Delta\Delta G$, in kcal/mol, from RosettaInterface Computational Mutagenesis for Clusters that Pass the Mutagenesis Filter and the Structural Superposition Elimination Procedure. Only EGFR mutations that lead to binding loss experimentally are listed. $\Delta\Delta G > 1.0$ indicates a hit (bold). The + or − signs, respectively, signify that the mutation resulted in partial or large binding loss in experiments (Chao et al., 2004). $\Delta\Delta G > 10$ usually indicates a severe steric clash between side chains. No distinction is made between + and − mutations in the definition of a hit.

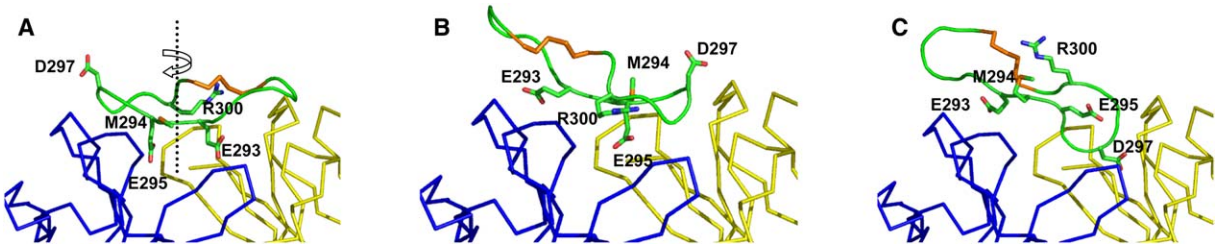

Figure 4. mAb 806-Peptide Interface in the Preliminary Models

(A–C) The mAb 806-peptide interface in the preliminary models. (A) Model 1; (B) model 2; (C) model 3. The mAb 806 light and heavy chains are colored yellow and blue, respectively. The disulfide bond between EGFR residues 287 and 302 in the peptide epitope is shown in orange. In all three models, E293 makes energetically important contacts at the interface with either CDR H3 (model 1) or CDR H2 (models 2 and 3). Model 2 may be obtained from model 1 by rotating the peptide about the axis shown in (A). In models 1 and 2, M294, E295, and R300 point toward the interface, and D297 points away from the interface. The reverse is true for model 3.

monomer. The superposition with the hypothetical extended monomer has only minor clashes, in contrast with the interpenetration seen for the tethered monomer. The minor clashes with the hypothetical extended monomer can be relieved by local motion of the dimerization arm and residues 303–310 distal to the peptide epitope while keeping the mAb 806-peptide orientation constant.

Experimental Design to Discriminate between the Three Models

Given the completely different realizations of the intermediate state from models 1 and 2 and model 3, we performed further computational and experimental mutagenesis to discriminate between the models. We computationally mutated each peptide residue in the three mAb 806-peptide models to each one of the other 17 amino acids (excluding Cys and Pro). For several aromatic substitutions, we predicted a loss of binding (large positive $\Delta\Delta G$) due to steric clashes of the mutant side chains with the antibody CDRs. The interesting aspect of these predictions was the differences in the trends exhibited by models 1 and 2 compared to model 3 for aromatic substitutions of residues 293–301 (Table 4). In general, the majority of the aromatic substitutions were predicted to diminish binding in models 1 and 2, but were predicted to be neutral in model 3. We tested the predictions by performing a new set of site-directed mutagenesis experiments. We list the results of these experiments in Table 4, and compare them to the computational predictions for the three models. Aromatic substitutions of residues E293 and D297 result in binding loss, while the remaining aromatic mutations are neutral toward mAb 806 binding. Only the predictions for model 3 are consistent with this trend, while the predictions for models 1 and 2 are mostly inconsistent. Based on these observations, we propose that the relative orientation of mAb 806 and the peptide epitope in the mAb 806-EGFR complex is well approximated by model 3. This also implies that the intermediate EGFR state recognized by the antibody is closer to the hypothetical extended monomer than the tethered monomer state. The coordinates of this model have been deposited in the Protein Data Bank with code 2EXP.

Characteristics of the Proposed Structure of the mAb 806-EGFR Complex

The proposed structure (Figure 6A) was created by structurally aligning model 3 and the hypothetical extended monomer state. These coordinates have been deposited in the PDB with code 2EXQ. The interface buries 1224 Å^2 of surface area between the peptide and the antibody, including 721 Å^2 of buried hydrophobic surface and 503 Å^2 of buried hydrophilic surface (calculated by using Naccess [Hubbard and Thornton, 1993]). The interface incorporates 16 antibody CDR residues and all peptide residues except C287, G288, A289, and R300. The mAb contacts include 1, 2, and 3 residues on CDRs L1, L2, and L3, respectively, and 3, 5, and 2 residues on CDRs H1, H2, and H3, respectively. There are seven hydrogen bonds across the interface: E293-Y50H, Y292-N56H, E295-R97H, K301-A92L, Y292-S54H, E296-A33H, and D297-Y91L (in decreasing order of calculated Rosetta

Table 3. Antibody Interface Residues

Peptide Side Chain	Model 1	Model 2	Model 3
E293	L: H50, Y91, W96; H: A95, F99	H: Y50, S52, S54, N56	H: Y50, S52, Y53, S54, N56
M294	H: D31A, F32	L: Y91, A92, F94	H: D31A, Y53
E295	L: W96; H: A33, N35, Y50, S52	L: Y91; H: A95	H: R97
E296	H: Y50, S52, Y53	L: N32, H50, Y91	L: Y91, W96; H: F32, A33
D297	H: N56, R58	—	L: Y49, H50, Y91; H: R97, F99
G298	L: F94; H: Y50, R58	H: G96, R97	H: R97
V299	—	—	L: N32, H50, Y91
R300	L: N32, Y91, A92, Q93	H: D31A, F32	—
K301	—	—	L: N32, Y91, A92

Antibody CDR Residues that Contact Peptide Residues in the Three Models. A contact is made when two side chains have at least one nonhydrogen atom pair within 4 Å of each other. Antibody residues follow Chothia numbering. The light (L) and heavy (H) chain CDR regions are residues 24–34, 50–56, 89–97 and 26−35, and 50−65 and 95–102.

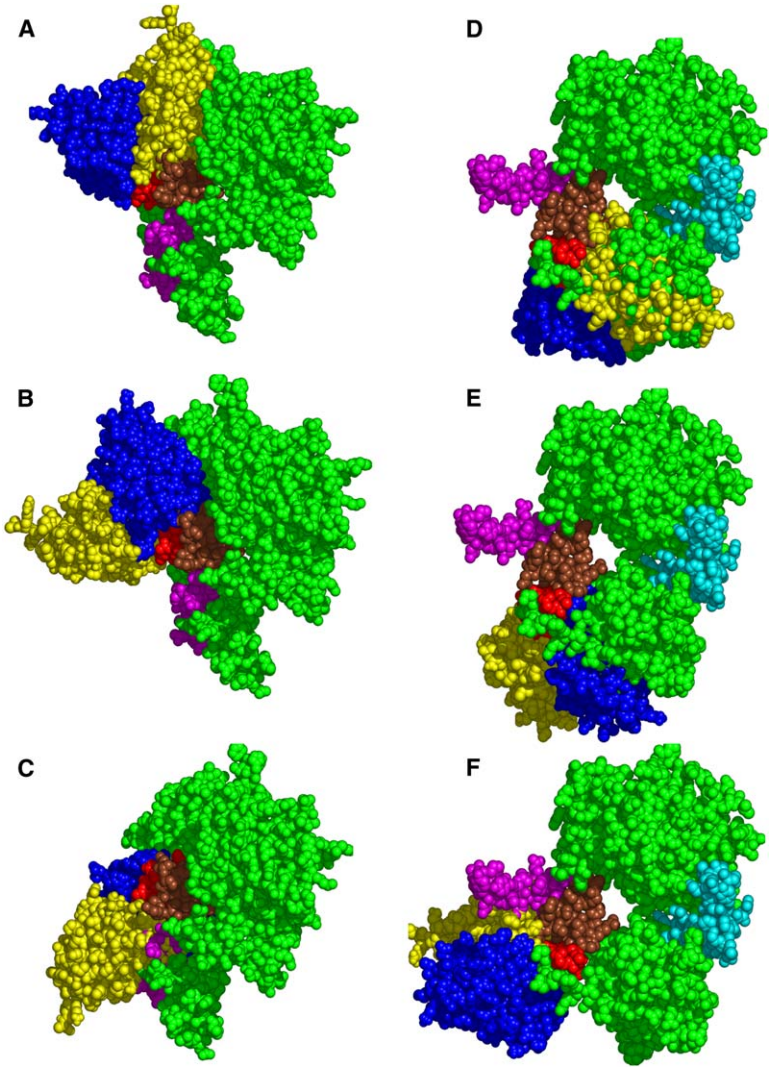

Figure 5. Structural Superposition of the Preliminary Models with EGFR Crystal Structures

(A–F) Structural superposition of models 1, 2, and 3 with the (A–C) tethered and the (D–F) hypothetical extended monomer conformation lacking domain IV. The EGFR dimerization arm is colored magenta, EGFR residues 261–286 (postdimerization arm segment) are brown, the mAb 806-peptide epitope is red, and all other EGFR residues are green. The EGF ligand (not shown in the tethered EGFR) is colored cyan. mAb 806 light and heavy chains are shown in yellow and blue, respectively. The final proposed model is (F). In the model, mAb 806 binds the hypothetical extended monomer with minor clashes of the antibody with the dimerization arm (magenta) and EGFR residues 303–310.

hydrogen bonding energy, where Y50H refers to the Tyr at residue position 50 of the heavy chain, etc.). Several residue pairs also make extensive hydrophobic contacts: D297-Y91L, D297-R97H, Y292-Y53H, E296-A33H, M294-Y53H, K301-A92L, D297-Y49L, and E293-N56H (in decreasing order of Rosetta van der Waals attractive energies). The intermolecular contacts that contribute the most to the total binding energy in the Rosetta calculations are E293-Y50H and D297-Y91L. The residue-level interactions as elucidated by the model suggest explanations for the energetics seen in the experimental mutagenesis. E293 mutations (to D, G, or K) cause binding loss due to the loss of a hydrogen bond to Y50H and the loss of hydrophobic contacts with CDR H2 residues (Figure 6B). The binding loss due to the D297Y mutation can be attributed to the steric repulsion between the mutant side chain and the aromatic side chains forming the antibody pocket (Figure 6C).

Discussion

The prediction of antibody-antigen complex structures is critical to rational biopharmaceutical design, as the location of neutralizing epitopes on cell surface receptors carries significant information regarding the receptor's structure/function relationships in normal and diseased physiology. A particular challenge in the prediction of antibody-antigen docking is the hypervariability of the energetically important antibody CDR loop H3. An advantage for prediction of antibody-antigen-docked structures over other complexes is the knowledge that CDR loops generally dominate antigen contacts; however, even so, in the latest CAPRI open challenge to compare computational docking algorithms, the statistics for successful antibody-antigen complex formation were no better than for other complex types, and, in some cases, they were significantly worse than average (Janin, 2005; Mendez et al., 2003, 2005; Vajda and Camacho, 2004; Wodak and Mendez, 2004). In this work, we have attempted to combine computational and experimental mutagenesis to discriminate false positives among candidate antibody-antigen-docked structures.

Due to the unavailability of the structure of the full-length intermediate EGFR conformation recognized by mAb 806, we split the structure prediction problem into two parts. First, we created an energetically favorable docking model of mAb 806 bound to the 16 residue EGFR epitope (Figure 4). Next, we inferred the global

Table 4. Mutagenesis Predictions to Discriminate and Validate Preliminary Models

Mutation	Model 1	Model 2	Model 3	Experimental Result
E293Y	0.8	**13.7**	**9.0**	—
E293F	−2.8	**12.0**	**8.1**	—
E293W	**3.4**	**14.0**	**6.4**	—
M294Y	2.2	4.2	**−0.4**	++
M294F	**−0.8**	1.9	**−0.2**	++
M294W	15.5	16.8	**0.8**	++
E295Y	48.1	**0.0**	**0.4**	++
E295F	36.6	**−0.7**	**0.4**	++
E295W	46.7	−1.6	**0.2**	++
E296W	**0.9**	2.5	−1.3	++
D297W	−0.03	−0.1	**31.0**	+
V299Y	**−0.3**	**−0.9**	**−0.8**	++
R300Y	17.7	17.9	**−0.4**	++
R300F	11.1	9.1	**−0.4**	++
R300W	9.2	8.7	**−0.4**	++
K301W	**0.0**	**−0.1**	1.4	++

Calculated $\Delta\Delta G$ Values for Aromatic Substitutions of Peptide Residues 293–301 for the Three Preliminary Models. Predictions consistent with experimental results are shown in bold. In the experimental results column, ++, +, and − indicate mutations that lead to zero, partial, and large binding loss, respectively. $\Delta\Delta G > 10$ usually indicates a severe steric clash between side chains.

structure of the complex (Figures 5F and 6) by superimposing this model with the EGFR ectodomain crystal structure and using known information regarding EGFR conformational change to predict the relative arrangement of the antibody and the EGFR. Accordingly, the proposed structure of the 806-EGFR complex (Figure 6A) has two important characteristics corresponding to two aspects of the structure prediction problem.

The first important characteristic is that if the energetically favorable antibody-peptide orientation seen in model 3 is to be permitted without steric hindrance from the remaining EGFR residues, the EGFR molecule has to adopt a conformation that is in the conformational vicinity of the hypothetical extended state rather than the tethered monomer. This conclusion arises because the dimerization arm residues clash with the antibody residues when model 3 is superimposed with the tethered monomer (Figure 5C). The final structure built by using the hypothetical extended monomer state of EGFR retains minor van der Waals bumps between the antibody and the EGFR that would need to be annealed by small domain motions of the EGFR. The biological implication of the predicted structure is that the steric hindrance created by the mAb 806 in the region of the dimer interface would prevent receptor dimerization and subsequent tyrosine kinase activation. We postulate this to be the structural origin of the molecular mechanism by which mAb 806 induces an antitumor effect. The hypothesis that receptor untethering is a prerequisite for mAb 806-EGFR binding is corroborated by several pieces of experimental evidence (Johns et al., 2004; Walker et al., 2004), and our model for the mAb 806-EGFR complex is consistent with these experimental observations. This mechanism also explains why mAb 806 only binds cancer cells, as (1) the epitope is always exposed in the de2-7 cancerous EGFR mutant (residues 6–273 deleted), and (2) improper glycosylation can affect the relative stability of the tethered state and shift the population toward an extended, untethered state.

The second important characteristic of the model corresponding to the antibody-epitope relative orientation is that it requires the "flipped out" peptide epitope backbone observed only in the tethered crystal structure (1NQL), and not in the dimer (1MOX or 1IVO) or the recently published structure of the tethered EGFR complexed with Cetuximab (1YY9) (Li et al., 2005). However, only the "flipped out" peptide conformation has the specific side chain orientations that correctly capture the differences in the effect of peptide residue mutations on mAb 806 binding. All EGFR residues essential for mAb 806 binding are contained in the peptide (Johns et al., 2004), but experimental alanine mutations at all positions except E293 and G298 are neutral toward mAb 806 binding (Chao et al., 2004). Only nonconservative mutations (D297Y, V299D, and K301E) indicate the energetic importance of other residues. Each of the preliminary models cleared the mutagenesis filter based on these data (Table 2). However, our new experimental results (Table 4) indicate that aromatic substitutions at positions 294–296 and 299–301 do not decrease binding affinity. Of the three models, only model 3 correctly predicts the majority of these neutral mutations. The difference in the predictions of model 3 compared to models 1 and 2 can be traced to a change in the directionality of residue E293 between the "flipped out" and "flipped in" peptide conformations (Figure 3).

In the "flipped in" conformation (models 1 and 2), the orientations of the C_α-C_β vectors of residues 293 and 297 differ by an average angle of 118°, indicating that the side chains point in nominally opposite directions. Thus, D297 has a significantly reduced chance of contacting antibody residues at the interface if E293 is at the interface and vice versa. However in the "flipped out" conformation, this angle is 9.5°, indicating that the side chains of E293 and D297 have a similar directionality and can occur at the interface simultaneously. Thus, model 3 captures the binding loss due to substitutions at both E293 and D297, while models 1 and 2 show agreement only with the E293 binding loss mutations. Similarly, in the "flipped in" conformation, M294 and R300 are oriented such that the side chains are buried at the interface if E293 is an interface residue. Consequently, in models 1 and 2, E293 binding loss mutations also imply binding loss due to M294 and R300 aromatic substitutions (Table 4). In the "flipped out" peptide conformation, the direction of residue E293 results in M294 and R300 pointing opposite of E293. Therefore, the "flipped out" side chain orientations in model 3 are consistent with both the E293 binding loss mutations and neutral aromatic substitutions of M294 and R300.

Since the model is dependent on the computational mutagenesis results, it is important to analyze the limitations of this technique and the mismatches between the final model and the experiment. In the original RosettaInterface work, Kortemme and Baker (2002) were able to successfully identify 79% of hotspots and 68% of neutral interface residues in a benchmark of 19 protein complexes. This is remarkable given that RosettaInterface models the residue mutations on a fixed protein backbone and does not account for binding loss due to local unfolding or accommodate mutations that might

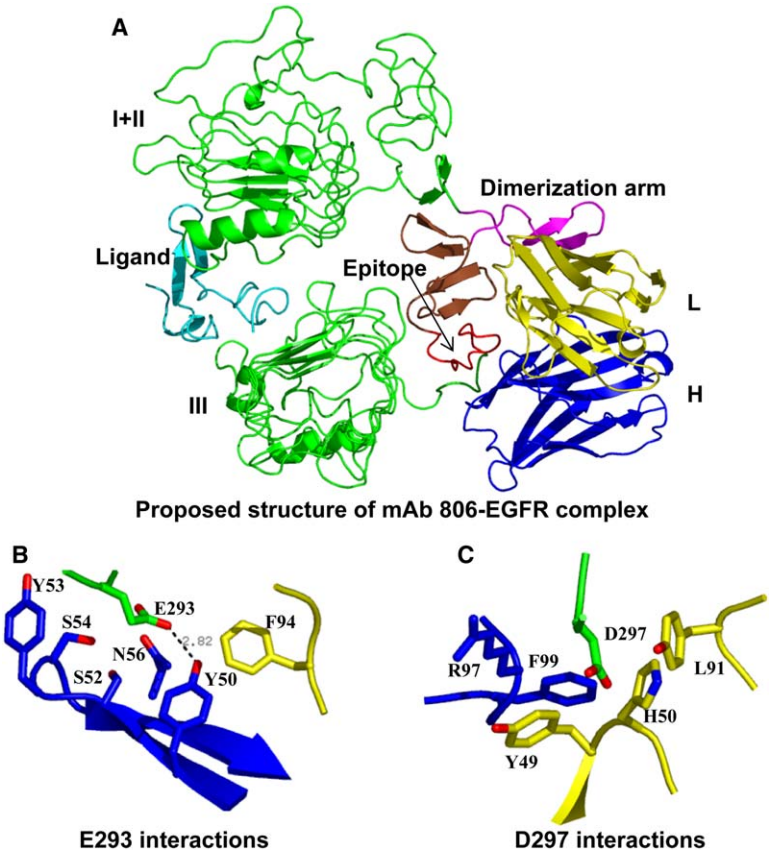

Figure 6. Proposed Structure for the mAb 806-EGFR Complex

(A) Proposed structure for the mAb 806-EGFR complex, created by structurally aligning model 3 with the hypothetical extended monomer conformation. The coloring scheme is identical to that in Figure 5.

(B) Interface interactions: EGFR residue E293 forms a hydrogen bond with CDR H2 side chain Y50 and makes several hydrophobic contacts, mostly with residues in the mAb 806 H2 loop.

(C) EGFR residue D297 is well packed in a pocket of residues from the mAb 806 L2, L3, and H3 CDRs.

become neutral via local backbone motion in real proteins. Overall, model 3 correctly predicts 33 of 40 (80%) mutations, including 14 of 16 (87%) new mutations checked experimentally *after* the model had been built. E296W ($\Delta\Delta G = -1.3$) and K301W ($\Delta\Delta G = +1.4$) are incorrectly predicted to be stabilizing and destabilizing, respectively, in model 3, as these mutations are neutral in experiments. Model 3 incorrectly predicts that mutations G298A, G298S, V299D, and K301E are neutral when a binding loss is seen in experiments. The side chains of V299 and K301 make interface contacts in model 3, but these interactions are either not accurately modeled during docking, or the energies of these interactions are not correctly revealed by computational mutagenesis. Glycine residues have no side chain and thus tolerate a relatively wide range of (ϕ,φ) values. RosettaInterface was parameterized by explicitly omitting mutations to or from glycine residues due to the possibility of backbone conformational change. In fact, four of the original ten binding loss mutations considered involve a change to or from glycine (E293G, G298A, G298D, G298S). In the final model, the E293 (ϕ,φ) coordinate of the peptide backbone is located on the Ramachandran map in an area favored by both glycine and nonglycine residues (data not shown). Therefore, the E293G mutation might not change the backbone conformation (the computational and experimental mutagenesis agree for this position). Also, a hit is obtained for the G298D mutation, but not for the G298A and G298S mutations (Table 2). The G298 (ϕ,φ) coordinate in model 3 is favored for glycine residues, but it is not permitted for nonglycine residues. The binding loss due to the G298A/S mutations might be caused by conformational change, explaining why the effects of these mutations are not captured by our simulations.

There might also be errors in the absolute values of $\Delta\Delta G$ due to general errors in the energy function. In the original RosettaInterface work, $\Delta\Delta G$ predictions for mutations involving charged residues were underestimated, even when the residues were qualitatively identified as hotspots. Likewise, in the present work, the predicted $\Delta\Delta G$ for the E293K mutation of 1.5 kcal/mol (Table 2) is less than the experimental result of 5.7 kcal/mol (Chao et al., 2004), probably due to the simplified treatment of the electrostatic interactions in the RosettaInterface model. Due to the lack of quantitative agreement, as suggested by Kortemme and Baker (2002), we use $\Delta\Delta G$ greater than or less than 1.0 kcal/mol as the criterion for labeling neutral and binding loss mutations. A further caveat for these predictions is that we are calculating $\Delta\Delta G$ values on a *model* of the docked complex, rather than a crystal structure. To overcome additional uncertainties, we oversampled the conformational space significantly and selected structures with agreement with the experimental mutagenesis data. The qualitative agreement of the computational and experimental mutagenesis on the models to an extent near that of the agreement of previous calculations on crystal structures is encouraging.

The docking results are dependent on the accuracy of the crystal structure conformations of the peptide epitope and the homology model for mAb 806. The disulfide

bond between the N- and C-terminal cysteines stabilizes the peptide backbone conformation and ensures that a large conformational change (unfolding) is starkly impossible. In addition, our use of five peptide conformations from three independently determined crystal structures is aimed at capturing the variability in the backbone conformation seen especially for residues 292–298. For mAb 806, the backbone conformations of all CDR loops except H3 are identical to the respective conformations of antibody M02/05/01, which shares identical CDR canonical classifications. The backbone conformation of CDR H3 is similar to the conformation of the homologous CDR H3 in antibody PC282. Potential errors in the homology model are restricted to the positioning of CDR H3 with respect to the antibody framework. While this is cause for some concern, the short length of H3 is a mitigating factor. Analysis of the antigen bound conformations of two antibodies with 7 residue CDR H3s (HYHEL-5 [PDB 1BQL] and D44.1 [PDB 1MLC]) in a protein-protein docking benchmark (Chen et al., 2003) indicates that the bulk of the interface contacts in these complexes is contributed by CDRs L3 and H2, and that CDR H3 plays a diminished role in antigen recognition. Similarly, in the final model, most of the contacts are mediated by residues from CDRs L3 and H2, and only two CDR H3 residues (R97 and F99) contact the antigen.

Any backbone errors in either the peptide or mAb 806 might impact the energetics of binding for individual peptide residues and result in incorrect predictions for the mAb 806-peptide orientation. However, the post-docking mutagenesis results provide strong support for the model. It is unlikely that the predicted result of the mutations would be consistent with greater than 80% of experimental results in an incorrect model. Several matches with experimental mutagenesis measurements were made *after* the predictions of the candidate models, and the new experimental information was not used in generating the model. Producing robust docking solutions in spite of errors in the backbone conformation is not without precedent. In fact, RosettaDock predicted the cohesin-dockerin CAPRI target (dockerin homology model with 52% sequence identity) with an 1.17 Å interface rmsd and 42% of the native contacts despite significant deviations in the backbone of the homology-modeled dockerin (Daily et al., 2005).

The mAb 806-EGFR system has served as a test case for a novel, to our knowledge, approach that combines protein docking with computational and experimental mutagenesis to model protein complexes. Computational mutagenesis is more specific and informative than using interface contacts to filter docking solutions. For instance, the fact that M294 is an interface residue in all of the models (Table 3) does not shed light on the effect of aromatic substitutions at this position and therefore has no value in discriminating between the three models. However, with the computational mutagenesis calculation, this residue's experimental results clearly support only one model (Table 4). As further evidence of the predictive power added by computational mutagenesis, we have used the protocol developed in this mAb 806-EGFR study to predict the structure for CAPRI target 21 (Sir1-Orc1 complex [Hou et al., 2005]) for which 7 neutral mutations and 7 binding loss mutations had been experimentally characterized (Bose et al., 2004; Gardner et al., 1999). The best predictions by any of the CAPRI participants were of "medium" accuracy ("two-star," i.e., [a] 30%–50% native residue-residue contacts and ligand or interface rmsd < 5 Å, or [b] greater than 50% native residue-residue contacts and ligand or interface rmsd > 1 Å [Mendez et al., 2005]). We were one of only three groups to produce such a model out of 37 CAPRI participants. HADDOCK (Dominguez et al., 2003), which also incorporates biochemical information, achieved a medium-accuracy prediction as well. In our prediction of this complex, computational mutagenesis correctly identified three of the seven binding loss mutations (Y489S, R493G, A505D) in our best model and in the subsequently released crystal structure (Hou et al., 2005). These results are a testament to the quality of RosettaDock's side chain refinement protocol and the utility of combining docking with computational mutagenesis for blind docking predictions.

In summary, we have proposed a model of the structure of the mAb 806-EGFR complex by combining a diverse set of computational tools to overcome several modeling challenges. These challenges include antibody homology modeling, CDR loop uncertainty, a large and flexible antigen, and subsequent verification of the predicted model. The model will provide a working hypothesis for further analysis and engineering of mAb 806 for use in cancer therapy. This case study suggests that the field of protein structure prediction now offers significant modeling capability to approach complex problems. In addition, we closely combined docking with experimental mutagenesis information to solve the mAb 806-EGFR complex. Mutagenesis data for the interaction of protein binding partners are quite frequently available or easily obtainable, but the complex structure is often unknown and less easily obtained. There is vigorous interest in the development of therapeutic antibodies against diverse indications such as cancer, inflammation, autoimmune disorders, and infectious diseases. As both computational docking and mutagenesis become more accurate with the inclusion of backbone flexibility and sophisticated treatment of the electrostatic interactions, our approach could be extremely useful for integrating experimental data with docking studies for this important class of protein-protein interactions.

Experimental Procedures

Starting Structures

mAb 806

The sequence (Old et al., 2002) of the mAb 806 Fv is as follows (the CDR regions are in bold). Light chain: DILMTQSPSSMSVSLGDTV SITC**HSSQDINSNIG**WLQQRPGKSFKGLIY**HGTNLDD**EVPSRFSG SGSGADYSLTISSLESEDFADYYC**VQYAQFPWT**FGGGTKLEIKRA; heavy chain: DVQLQESGPSLVKPSQSLSLTCTVT**GYSITSDFAWN** WIRQFPGNKLEWMG**YISYSGNTRYNPSLKS**RISITRDTSKNQFFL QLNSVTIEDTATYYCVT**AGRGFPY**WGQGTLVTVS. A single-chain variable fragment homology model for antibody 806 was built by using the Web Antibody Modeling (Whitelegg and Rees, 2000) (WAM) server. This method finds the closest antibody homolog for the light and heavy Fv region and then creates the conformations of the CDR loops from canonical classes by using a sequence-based algorithm with empirical rules. The rules typically fail for the CDR H3 loop, and, in this case, its configuration is determined with a combined database/conformational search.

EGFR Peptide

The five backbone conformations of the peptide epitope (EGFR residues 287–302) are taken from the tethered monomer (1NQL, resolution 2.8 Å) and the untethered homodimer crystal structures (1IVO, resolution 3.3 Å; 1MOX, resolution 2.5 Å). Side chain coordinates are missing for several peptide residues in crystal structures of the dimer 1MOX (290, 295, and 299–301 in chain A; 290, 296, 300, and 301 in chain B) and the tethered monomer 1NQL (290, 293–297). These are reconstructed during docking by using a backbone-dependent rotamer library (Dunbrack and Cohen, 1997).

Docking

We use the RosettaDock algorithm for the mAb 806-peptide docking simulations. We developed RosettaDock to predict a protein complex structure from its unbound monomer components (Gray et al., 2003b). Independent simulations are carried out in parallel, and the resulting "decoys" are ranked by using a scoring function dominated by van der Waals interactions, an implicit solvation model (Lazaridis and Karplus, 1999), and an orientation-dependent hydrogen bonding potential (Kortemme et al., 2003). We generate 10^5 docking decoys in a global docking search by sampling all orientations of the peptide in the vicinity of the antibody CDRs, and we use an alignment score to favor peptide contacts with the CDRs (Gray et al., 2003a). The RosettaDock protocol is available free for academic use at http://graylab.jhu.edu.

Computational Mutagenesis

The RosettaInterface computational mutagenesis approach (Kortemme and Baker, 2002) was used to estimate the change in the binding free energy ($\Delta\Delta G$) for individual residue mutations. This algorithm estimates $\Delta\Delta G$ by modeling the residue mutation in both the protein complex and the uncomplexed monomers using a backbone-dependent rotamer library. The energy function is similar to the RosettaDock energy function and incorporates the solute-solvent entropy and the solvent reorganization entropy via the Lazaridis-Karplus implicit solvation model, but neglects the protein configurational entropy. The RosettaInterface computational mutagenesis server can be accessed at http://robetta.bakerlab.org.

Structural Superposition to Eliminate Unphysical mAb 806-Peptide Orientations

A comparative analysis of the tethered (1NQL) and untethered (1MOX, 1IVO) EGFR crystal structures suggests that five residue segments stay rigid, with relative motion amongst these segments during the conformational change. The fragments, numbered 1–5, are as follows: residues 3–240, containing all of domain I and parts of domain II; residues 241–260, containing the dimerization arm; residues 261–286 that occur prior to the mAb 806 epitope; residues 287–302, containing the epitope; and residues 310–512, containing domain III. We calculated the separation between the centroids of the peptide epitope and each of the other segments as defined above individually for the tethered and the untethered states. Then, we calculated the change in this centroid-centroid separation between the tethered and the untethered states. This change in the centroid-centroid separation between the peptide epitope and fragments 1, 2, 3, and 5 are 7.1 Å, 1.5 Å, 0.4 Å, and −9.4 Å, respectively. The magnitude of this change indicates the extent to which the segments move relative to the peptide epitope 287–302 due to the conformational transition. Since the change in the centroid-centroid separation between fragment 3 (residues 261–286) and the epitope due to the conformational change is negligible (<0.5 Å), this implies that fragment 261–302 moves as a rigid body.

The design of the structural superposition filter to eliminate physically unrealistic mAb 806-peptide orientations is as follows. For a given docking decoy, we perform a least-squares structural superposition of the C_α atoms of the peptide residues in the decoy and the arbitrary tethered EGFR crystal structure (1NQL) by using PyMol (DeLano, 2002). Since mAb 806 does not bind the tethered EGFR in experiments, the superposition could legitimately result in clashes of mAb 806 residues with the nonpeptide EGFR residues. These clashes may be "minor" van der Waals bumps involving less than 5 residues from each protein, or they may result in segment-segment interpenetration involving more than 25 residues from each protein. If the clashes are between mAb 806 and EGFR residue fragments that move relative to the peptide and away from it, it is theoretically possible for the EGFR residue fragments to rearrange themselves into an intermediate state to permit the mAb 806-peptide orientation without any clashes. On the other hand, the clashes might involve residue fragments that do not appear to move relative to the peptide or move toward it. In this case, the clashes are likely to persist during the conformational change, and such mAb 806-peptide orientations are physically unrealistic. Therefore, we eliminated from any further consideration all decoys that resulted in complete interpenetration of mAb 806 and residue segment 261–286, since this segment does not appear to move relative to the peptide epitope between the tethered and untethered EGFR. We also eliminated those decoys that displayed large interpenetration of more than 25 residues with either segment 3–240 or segment 310–512. By the same token, we retained decoys that had minor clashes of less than 5 residues with any of the five residue segments. We also produced animations of the EGFR conformational change by using the Morph Server (Echols et al., 2003) and the Elastic Network Interpolation method (Kim et al., 2003) to confirm our above-described elimination strategy. While these interpolations do not necessarily capture the biophysical pathway of the EGFR conformational change, they offer useful guidelines for interpreting the results of the structural superposition.

Site-Directed Mutagenesis Experiments

Site-directed mutagenesis of EGFR and testing were performed as described previously (Chao et al., 2004). Briefly, amino acid mutations of EGFR fragment 273–621 in the yeast surface display plasmid pCT302 were introduced by using mutagenic PCR primers. The mutant plasmids were transformed into yeast strain EBY100 and were galactose induced. Mutants on the surface of yeast were tested for binding at 75 nM mAb 806 by using flow cytometry. mAb 806 was generously provided by the Ludwig Institute for Cancer Research.

Acknowledgments

A.S. acknowledges support from the Chemical Therapeutics Working Group at the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins. This work was funded by National Institutes of Health (NIH) R01-CA96504 for K.D.W. and NIH K01-HG02316 for J.J.G. G.C. is the recipient of a National Defense Science and Engineering Graduate Fellowship. mAb 806 was generously provided by Achim Jungbluth of the Ludwig Institute for Cancer Research. J.J.G. and A.S. acknowledge Greg Chirikjian and Jay Jeong for assistance with the Elastic Network Interpolation method for analyzing the pathway of the EGFR conformational change. Thanks to Dan Leahy and Greg Bowman for discussion and review of our manuscript.

Received: August 26, 2005
Revised: November 9, 2005
Accepted: November 16, 2005
Published online: March 14, 2006

References

Al-Lazikani, B., Lesk, A.M., and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. J. Mol. Biol. *273*, 927–948.

Aloy, P., Querol, E., Aviles, F.X., and Sternberg, M.J.E. (2001). Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. J. Mol. Biol. *311*, 395–408.

Berchanski, A., Shapira, B., and Eisenstein, M. (2004). Hydrophobic complementarity in protein-protein docking. Proteins *56*, 130–142.

Bose, M.E., McConnell, K.H., Gardner-Aukema, K.A., Muller, U., Weinreich, M., Keck, J.L., and Fox, C.A. (2004). The origin recognition complex and Sir4 protein recruit Sir1p to yeast silent chromatin through independent interactions requiring a common Sir1p domain. Mol. Cell. Biol. *24*, 774–786.

Bracci, L., Pini, A., Bernini, A., Lelli, B., Ricci, C., Scarselli, M., Niccolai, N., and Neri, P. (2003). Biochemical filtering of a protein-protein

docking simulation identifies the structure of a complex between a recombinant antibody fragment and α-bungarotoxin. Biochem. J. *371*, 423–427.

Camacho, C.J., and Gatchell, D.W. (2003). Successful discrimination of protein interactions. Proteins *52*, 92–97.

Chao, G., Cochran, J.R., and Dane Wittrup, K. (2004). Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display. J. Mol. Biol. *342*, 539–550.

Chen, R., Mintseris, J., Janin, J., and Weng, Z.P. (2003). A protein-protein docking benchmark. Proteins *52*, 88–91.

Cho, H.S., Mason, K., Ramyar, K.X., Stanley, A.M., Gabelli, S.B., Denney, D.W., and Leahy, D.J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature *421*, 756–760.

Chothia, C., and Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. *196*, 901–917.

Comeau, S.R., Gatchell, D.W., Vajda, S., and Camacho, C.J. (2004). ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res. *32*, W96–W99.

Daily, M.D., Masica, D., Sivasubramanian, A., Somarouthu, S., and Gray, J.J. (2005). CAPRI rounds 3–5 reveal promising successes and future challenges for RosettaDock. Proteins *60*, 181–186.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Dominguez, C., Boelens, R., and Bonvin, A.M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. *125*, 1731–1737.

Dunbrack, R.L., Jr., and Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. *6*, 1661–1681.

Echols, N., Milburn, D., and Gerstein, M. (2003). MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res. *31*, 478–482.

Ferguson, K.M., Berger, M.B., Mendrola, J.M., Cho, H.S., Leahy, D.J., and Lemmon, M.A. (2003). EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. Mol. Cell *11*, 507–517.

Fernandez-Recio, J., Totrov, M., and Abagyan, R. (2003). ICM-DISCO docking by global energy optimization with fully flexible side-chains. Proteins *52*, 113–117.

Franklin, M.C., Carey, K.D., Vajdos, F.F., Leahy, D.J., de Vos, A.M., and Sliwkowski, M.X. (2004). Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. Cancer Cell *5*, 317–328.

Gardner, K.A., Rine, J., and Fox, C.A. (1999). A region of the Sir1 protein dedicated to recognition of a silencer and required for interaction with the Orc1 protein in *Saccharomyces cerevisiae*. Genetics *151*, 31–44.

Garrett, T.P.J., McKern, N.M., Lou, M.Z., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Zhu, H.J., Walker, F., Frenkel, M.J., Hoyne, P.A., et al. (2002). Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. Cell *110*, 763–773.

Gray, J.J., Moughon, S.E., Kortemme, T., Schueler-Furman, O., Misura, K.M., Morozov, A.V., and Baker, D. (2003a). Protein-protein docking predictions for the CAPRI experiment. Proteins *52*, 118–122.

Gray, J.J., Moughon, S.E., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., and Baker, D. (2003b). Protein-protein docking with simultaneous optimization of rigid body displacement and side-chain conformations. J. Mol. Biol. *331*, 281–299.

Harari, P.M. (2004). Epidermal growth factor receptor inhibition strategies in oncology. Endocr. Relat. Cancer *11*, 689–708.

Herzyk, P., and Hubbard, R.E. (1998). Using experimental information to produce a model of the transmembrane domain of the ion channel phospholamban. Biophys. J. *74*, 1203–1214.

Hou, Z., Bernstein, D.A., Fox, C.A., and Keck, J.L. (2005). Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proc. Natl. Acad. Sci. USA *102*, 8489–8494.

Hubbard, S.J., and Thornton, J.M. (1993). NACCESS (London: Department of Biochemistry and Molecular Biology, Univ. College).

Janin, J. (2005). Assessing predictions of protein-protein interaction: the CAPRI experiment. Protein Sci. *14*, 278–283.

Johns, T.G., Stockert, E., Ritter, G., Jungbluth, A.A., Huang, H.J.S., Cavenee, W.K., Smyth, F.E., Hall, C.M., Watson, N., Nice, E.C., et al. (2002). Novel monoclonal antibody specific for the DE2–7 epidermal growth factor receptor (EGFR) that also recognizes the EGFR expressed in cells containing amplification of the EGFR gene. Int. J. Cancer *98*, 398–408.

Johns, T.G., Adams, T.E., Cochran, J.R., Hall, N.E., Hoyne, P.A., Olsen, M.J., Kim, Y.S., Rothacker, J., Nice, E.C., Walker, F., et al. (2004). Identification of the epitope for the epidermal growth factor receptor-specific monoclonal antibody 806 reveals that it preferentially recognizes an untethered form of the receptor. J. Biol. Chem. *279*, 30375–30384.

Johns, T.G., Mellman, I., Cartwright, G.A., Ritter, G., Old, L.J., Burgess, A.W., and Scott, A.M. (2005). The antitumor monoclonal antibody 806 recognizes a high-mannose form of the EGF receptor that reaches the cell surface when cells over-express the receptor. FASEB J. *19*, 780–782.

Kim, M.K., Li, W., Shapiro, B.A., and Chirikjian, G.S. (2003). A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA. J. Biomol. Struct. Dyn. *21*, 395–405.

Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. Proc. Natl. Acad. Sci. USA *99*, 14116–14121.

Kortemme, T., Morozov, A.V., and Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J. Mol. Biol. *326*, 1239–1259.

Lazaridis, T., and Karplus, M. (1999). Effective energy function for proteins in solution. Proteins *35*, 133–152.

Li, L., Chen, R., and Weng, Z.P. (2003). RDOCK: refinement of rigid-body protein docking predictions. Proteins *53*, 693–707.

Li, S., Schmitz, K.R., Jeffrey, P.D., Wiltzius, J.J., Kussie, P., and Ferguson, K.M. (2005). Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. Cancer Cell *7*, 301–311.

Luwor, R.B., Johns, T.G., Murone, C., Huang, H.J.S., Cavenee, W.K., Ritter, G., Old, L.J., Burgess, A.W., and Scott, A.M. (2001). Monoclonal antibody 806 inhibits the growth of tumor xenografts expressing either the de2–7 or amplified epidermal growth factor receptor (EGFR) but not wild-type EGFR. Cancer Res. *61*, 5355–5361.

Mandell, J.G., Roberts, V.A., Pique, M.E., Kotlovyi, V., Mitchell, J.C., Nelson, E., Tsigelny, I., and Ten Eyck, L.F. (2001). Protein docking using continuum electrostatics and geometric fit. Protein Eng. *14*, 105–113.

Massova, I., and Kollman, P.A. (1999). Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. J. Am. Chem. Soc. *121*, 8133–8143.

Mendez, R., Leplae, R., De Maria, L., and Wodak, S.J. (2003). Assessment of blind predictions of protein-protein interactions: current status of docking methods. Proteins *52*, 51–67.

Mendez, R., Leplae, R., Lensink, M.F., and Wodak, S.J. (2005). Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins *60*, 150–169.

Mishima, K., Johns, T.G., Luwor, R.B., Scott, A.M., Stockert, E., Jungbluth, A.A., Ji, X.D., Suvarna, P., Voland, J.R., Old, L.J., et al. (2001). Growth suppression of intracranial xenografted glioblastomas overexpressing mutant epidermal growth factor receptors by systemic administration of monoclonal antibody (mAb) *806*, a novel monoclonal antibody directed to the receptor. Cancer Res. *61*, 5349–5354.

Nanda, V., and DeGrado, W.F. (2005). Automated use of mutagenesis data in structure prediction. Proteins *59*, 454–466.

Norledge, B.V., Petrovan, R.J., Ruf, W., and Olson, A.J. (2003). The tissue factor/factor VIIa/factor Xa complex: a model built by docking and site-directed mutagenesis. Proteins *53*, 640–648.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. Cell *110*, 775–787.

Old, L.J., Johns, T.G., Panousis, C., Scott, A.M., Renner, C., Ritter, G., Jungbluth, A., Stockert, E., Collins, P., Cavanee, W.K., et al. (2002). PCT International Patent WO/02/092771.

Onrust, R., Herzmark, P., Chi, P., Garcia, P.D., Lichtarge, O., Kingsley, C., and Bourne, H.R. (1997). Receptor and $\beta\gamma$ binding sites in the α subunit of the retinal G protein transducin. Science *275*, 381–384.

Ritchie, D.W., and Kemp, G.J.L. (2000). Protein docking using spherical polar Fourier correlations. Proteins *39*, 178–194.

Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., Nussinov, R., and Wolfson, H.J. (2003). Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. Proteins *52*, 107–112.

Smith, G.R., Sternberg, M.J.E., and Bates, P.A. (2005). The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. J. Mol. Biol. *347*, 1077–1101.

Sridhar, S.S., Seymour, L., and Shepherd, F.A. (2003). Inhibitors of epidermal-growth-factor receptors: a review of clinical research with a focus on non-small-cell lung cancer. Lancet Oncol. *4*, 397–406.

Vajda, S., and Camacho, C.J. (2004). Protein-protein docking: is the glass half-full or half-empty? Trends Biotechnol. *22*, 110–116.

van Dijk, A.D., Boelens, R., and Bonvin, A.M. (2005). Data-driven docking for the study of biomolecular complexes. FEBS J. *272*, 293–312.

Walker, F., Orchard, S.G., Jorissen, R.N., Hall, N.E., Zhang, H.H., Hoyne, P.A., Adams, T.E., Johns, T.G., Ward, C., Garrett, T.P.J., et al. (2004). CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor. J. Biol. Chem. *279*, 22387–22398.

Wang, C., Schueler-Furman, O., and Baker, D. (2005). Improved side-chain modeling for protein-protein docking. Protein Sci. *14*, 1328–1339.

Whitelegg, N.R., and Rees, A.R. (2000). WAM: an improved algorithm for modelling antibodies on the WEB. Protein Eng. *13*, 819–824.

Wodak, S.J., and Mendez, R. (2004). Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. Curr. Opin. Struct. Biol. *14*, 242–249.

Yarden, Y., and Sliwkowski, M.X. (2001). Untangling the ErbB signalling network. Nat. Rev. Mol. Cell Biol. *2*, 127–137.

Zacharias, M. (2003). Protein-protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci. *12*, 1271–1282.

Accession Numbers

The atomic coordinates of the final mAb 806-EGFR peptide model and its structural superposition with the EGFR hypothetical extended monomer have been deposited in the Protein Data Bank with codes 2EXP and 2EXQ, respectively.